Title

authors
CSAIL, MIT
authors@mit.edu

Label ($n = 66,470$)	Vulnerable	Not vulnerable
`TOD`	2,360 (4%)	64,110 (96%)
`IntUn`	4,409 (7%)	62,061 (93%)
`IntOv`	18,544 (29%)	47,928 (71%)
`StateChange`	1,266 (2%)	65,204 (98%)

Table 1: Labels summary

Abstract

abstract goes here.

1 Introduction

blah blah

2 Background

ML for programs. background

Graph neural networks. some more background

3 Approach

talk about our approach

4 Method

5 Experiments

In this work, we investigate the following research questions

- How
- How
- Do features selected by models trained on complex representations provide a rich qualitative understanding?

We built a balanced set of roughly 66k functions and used blah blah

The dataset pertaining to each label was split into a training and test set in the ratio 66%-33%. The dataset for each label was balanced - with roughly the same number of positive and negative samples, and stratified to ensure that the distribution of function types was similar in the train and test set.

Our models

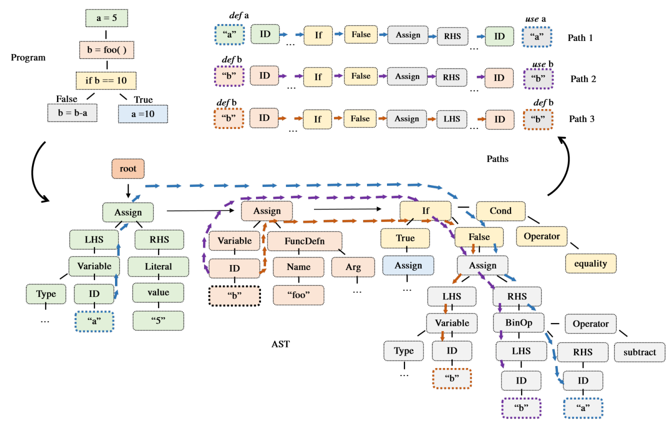

Figure 1: `Use & Define` Paths - An example. Paths to the expression $b = b - a$ are shown. Statements and their corresponding sub-trees have been shaded in the same color.

- **Model1, `Tr.`** Oh boy.
- **Model2, `Node.`** We use

For models trained on our representations, we also analyze the highest contributing features. These two analyses help answer the three research questions we pose.

6 Results

The test-set classification accuracies of the different models we evaluated are tabulated in Table 2.

7 Related Work

Recent work have focused on .

8 Conclusion

This work presents an important first step in

Label	N	Tr	Node	T	UD	UD+T	U	U+T	U+D+UD	U+D+UD+T
Tot # features		**2**	**44**	**79**	**8,232**	**8,311**	**14,940**	**15,019**	**23,172**	**23,251**
Avg # features trained on		**2**	**30**	**51**	**1,012**	**1,161**	**1,763**	**1,704**	**2,775**	**2,813**
TOD	**4,594**	0.47	0.72	0.70	0.78	0.80	0.76	0.79	0.80	0.80
StateChange	**2,440**	0.62	0.75	0.75	0.75	0.77	0.77	0.79	0.79	0.82
IntOv	**28,634**	0.66	0.78	0.81	0.82	0.84	0.84	0.86	0.85	0.87
IntUn	**6,676**	0.64	0.79	0.85	0.85	0.86	0.85	0.87	0.87	0.88

Table 2: Results summary. Test-set classification accuracies for ablation models.